

MOLECULAR STRUCTURE IN RELATION TO BIOLOGY AND MEDICINE

L. PAULING

California Institute of Technology, Pasadena

THE molecules that compose the body of a human being may be conveniently divided into two classes: small molecules and large molecules. Small molecules are molecules containing 10 or 20 or perhaps 100 atoms; examples are glucose, acetylcholine, glycine and other amino acids, thiamine and other vitamins. Large molecules are molecules containing hundreds or thousands or tens of thousands of atoms; examples are the proteins and the nucleic acids. To understand the human body in health and in disease we need to know the structure of the small molecules and the large molecules. The large molecules are especially important because it is they that carry biological specificity: their structural differences determine the differences between species of living organisms and also the differences between individuals of the same species.

During the past 40 years great progress has been made in the determination of the precise structures of small molecules. The work has been done by the use of several methods, of which the study of gas molecules by electron diffraction and of crystals by X-ray diffraction are the most important; much information is also now being provided by microwave spectroscopy and nuclear-spin and electron-spin magnetic resonance. Outstanding among the X-ray studies are the astounding achievements of Mrs. Dorothy Hodgkin and her co-workers, in determining the structures of penicillin and vitamin B₁₂.

During the recent decades there has also been developed a

powerful theory of atomic and molecular structure. This theory and the detailed experimental information about the structure of simple molecules and crystals have permitted chemical structure theory to be greatly extended and refined. I think that it now encompasses all of the significant structural principles necessary for the understanding of large molecules as well as small ones, and that no important new structural features will be discovered in the course of the structure determinations of proteins and nucleic acids that we expect to be made during the next decade or two.

Moreover, I think that the nature of the forces responsible for intermolecular interactions is now well understood, and that these forces—London dispersion force of van der Waals attraction, the force of van der Waals repulsion, the formation of hydrogen bonds, the interaction of electrically charged groups, the formation of fractional covalent bonds—can be confidently analysed as the basis of the characteristic intermolecular interactions leading to biological specificity, such as the interactions of antibodies and antigens.

I believe that biological specificity in general results from a detailed complementarity in structure of interacting molecules. There exists an overwhelming mass of evidence that the specific combining power of an antibody molecule for its homologous antigen molecule results from a complementarity in structure that permits the co-operation of several weak interactions that separately would not produce a significant bond between the molecules. This evidence has been provided by the work of Landsteiner with antigens containing haptenic groups with known chemical structure and by later studies along the same lines (Campbell, Pressman, Haurowitz). The combining powers of antibody molecules with haptens related in structure to the haptenic group of the immunizing antigen are found to change in the ways predicted for the interaction energy of the haptens with an antibody-combining region closely complementary to the original haptenic group. The replacement of an atom or radical

by a larger atom or radical causes a decrease in combining power, attributed to van der Waals repulsion (steric hindrance); replacement of a radical by a smaller one or by one of equal size with decreased polarizability (decreased power of van der Waals attraction), decreased electric charge, or decreased power of hydrogen-bond formation causes a decrease in combining power; replacement by one of approximately equal size and polarizability (such as a methyl group by a chlorine atom), even with much different chemical properties, results in no change in combining power.

The idea that the antigen or a fragment of it serves as the template against or about which a plastic material, the precursor of the antibody, is moulded through the operation of the forces of intermolecular attraction is an attractive one. We can understand the process of hardening of the antibody molecule in its complementary configuration through the formation of hydrogen bonds and the operation of other forces between the different parts of the folded polypeptide chain of the molecule. This postulated mechanism of formation of antibodies provides an explanation of many observations, such as the astounding versatility of the antibody-producing mechanism—the ability of an animal to manufacture specific antibodies against haptenic groups, such as the *p*-azobenzenearsonate ion, that probably have never constituted a part of the environment of the forebears of the animal. However, much remains to be discovered about the mechanisms of manufacture of proteins, and it may be found that these mechanisms are complex ones, involving a succession of steps. From the analysis of possible modes of operation of interatomic and intermolecular forces I have reached the conclusion that every step involving specificity will be found to depend for its specificity on a detailed complementariness in structure of the interacting molecules.

The importance of even the smallest structural details of the large molecules in the human body can be illustrated by the

discussion of the abnormal haemoglobins in relation to the hereditary haemoglobinopathies. It is now ten years since sickle cell anaemia was recognized as a molecular disease and the abnormal molecule responsible for it, haemoglobin S, was discovered (Pauling, Itano, Singer and Wells, 1949). During this decade many other abnormal forms of human haemoglobin have been discovered and many new diseases for which they are responsible have been described (see Pauling, 1955; Itano, 1956). The mechanism of the process of change in shape (sickling) of the erythrocytes of sickle cell anaemia patients has been recognized to be the formation of spindle-shaped tactoids (liquid crystals of the nematic type) of unoxygenated haemoglobin S. The formation of these crystals can be attributed to a self-complementariness in structure of the molecules of this protein. The self-complementariness is destroyed when the crystals combine with oxygen, and it is not shown by normal adult human haemoglobin (haemoglobin A).

A significant start has now been made on the determination of the difference in structure of haemoglobin S and haemoglobin A. Shortly after the discovery of haemoglobin S, Schroeder, Kay and Wells (1950) found that its amino acid composition is nearly the same as that of haemoglobin A. It was then found by Ingram (1958), by application of his powerful method of two-dimensional paper electrophoresis-chromatography to the enzyme-catalysed haemoglobin hydrolysates, that the difference in amino acid composition and sequence consists only in the replacement in each half-molecule of a glutamyl residue (in haemoglobin A) by a valyl residue (in haemoglobin S). (The haemoglobin molecule is shown to have a twofold symmetry axis, and hence to consist of two identical halves, by the X-ray diffraction pattern of the crystal.) There are about 600 amino acid residues in the haemoglobin molecule, and only two of the 600 are different in haemoglobin A and haemoglobin S; yet this small difference in structure is enough to cause the human beings who manufacture haemoglobin S to have a serious disease.

Something is now known about the location of the glutamyl-valyl replacement in the polypeptide chains. It was shown by Rhinesmith, Schroeder and Pauling (1957), by the use of Sanger's end-group method, that haemoglobin A contains two polypeptide chains of one kind (α chains, with N-terminal sequence val-leu) and two of a second kind (β chains, with N-terminal sequence val-his-leu). It has now been shown by Vinograd, Hutchinson and Schroeder (1959) that the glutamyl-valyl replacement occurs in the β chains; the α chains of haemoglobin A and haemoglobin S seem to be identical.

The method used by these investigators is an interesting one. Haemoglobin A labelled with ^{14}C was made by incubating human reticulocytes in blood to which L-leucine containing ^{14}C had been added. A solution containing labelled haemoglobin A and unlabelled haemoglobin S was brought to pH 5 for some hours—at this pH the α and β chains separate. The solution was brought back to pH 7, permitting the chains to recombine, and the two haemoglobins were separated by column chromatography. The N-terminal residues of the hybridized haemoglobin S were labelled with Sanger's reagent, the protein was partially hydrolysed, and the N-terminal peptides were isolated chromatographically and checked for radioactivity. The peptide DNP-val-leu, characteristic of α chains, was found to be strongly labelled with ^{14}C , and the peptide di-DNP-val-his-leu, characteristic of β chains, only weakly labelled. Hence, it is the β chains that are different in haemoglobin A and haemoglobin S.

Ingram (1959) has found the peptide val-his-leu-thr-pro-glu-glu-lys from haemoglobin A and val-his-leu-thr-pro-val-glu-lys from haemoglobin S, and has surmised (because the first three residues are the N-terminal set for the β chains) that the glutamyl-valyl replacement occurs in the sixth position from the N-terminus of the β chains. In haemoglobin C the same position is occupied by lysyl. He has also reported that for haemoglobins D_β and E the abnormalities are in the β chains and for D_α and I

they are in the α chains. Schwartz and co-workers (1957) have reported from studies of inheritance of S and G in families carrying both traits that different genes control the manufacture of haemoglobin S and haemoglobin G, and a similar conclusion about haemoglobin S and haemoglobin Hopkins-2 has been reached by Smith and Torbet (1958). It is possible that the abnormalities for haemoglobins G and Hopkins-2 are in the α chains, and that the synthesis of α chains and that of β chains are controlled by different genes.

Jones, Schroeder and Vinograd (1959) and Hunt (1959) have obtained evidence that human foetal haemoglobin, haemoglobin F, contains two α chains that are identical with those of haemoglobin A. Hence, an α -gene abnormality would be expected to cause the manufacture of two abnormal proteins, an abnormal foetal haemoglobin and an abnormal adult haemoglobin. This prediction has not yet been verified by observation.

It has been found by Jones and co-workers (1959) that haemoglobin H represents a new sort of molecular abnormality. Haemoglobin H was first reported in two children of Chinese descent; the investigators, Rigas, Koler and Osgood (1956), found no haemoglobin H in the red cells of the parents, whereas for other abnormal haemoglobins the trait is generally shown by one or both of the parents. The haemoglobin-H molecule was found by Jones and his co-workers to have four polypeptide chains with the same N-terminal sequence, val-his-leu. This observation suggested that the molecule consists of four normal β chains, and this hypothesis was then verified by a hybridization experiment with haemoglobin A labelled with ^{14}C and unlabelled haemoglobin H. In the course of this work a new haemoglobin, consisting of four sickle-cell β chains, was made; it is likely that at some future time this haemoglobin too will be found in Nature. The genetic abnormality that leads to the manufacture of haemoglobin H is apparently one that inhibits the synthesis of the α chains.

Thus, significant progress has been made in the study of the chemical structure of the abnormal haemoglobins, and yet we are still far from understanding the properties of the substances in terms of the structures of their molecules. It is almost certain that this understanding would not be achieved even though complete determinations were to be made of the amino acid sequences in the polypeptide chains of normal adult human haemoglobin and the various abnormal haemoglobins. Complete sequence studies made for insulin by Sanger and his collaborators have not led to an understanding of the physiological properties of this hormone. What is lacking as yet is knowledge about the detailed method of folding of the polypeptide chains and the configuration of the side chains; what is needed is the determination of the complete molecular structure of these proteins, and also of the proteins and other substances with which they interact. Despite the vigorous efforts of many investigators (among them are Perutz, Kendrew, Mrs. Hodgkin, Corey, Harker, and Bernal, and their collaborators) there has not yet been carried out the complete structure determination of any protein molecule. I estimate 1st March 1967 ± 2.5 years as the date when the announcement will be made that the first complete structure determination for a protein molecule, the determination by experiment (X-ray diffraction) of the relative positions in space of all of the atoms in the molecule, has been accomplished. The determination of the complete structure of a molecule of deoxyribonucleic acid will probably occur a few years later. Twenty-five years from now we shall probably know the complete structures of one hundred protein molecules and a few nucleic acid molecules. We shall then have a detailed understanding of the ways in which a few enzymes carry out their specific activities, the ways in which genes duplicate themselves and accomplish their individual tasks of precisely controlling the synthesis of protein molecules with well defined structures, the ways in which abnormal molecules give rise to the manifestations of the diseases that they cause, the ways in which

drugs and other physiologically active substances achieve their effects. When this time comes, medicine will have made a significant start in its transformation from macroscopic and cellular medicine to molecular medicine.

REFERENCES

- HUNT, J. A. (1959). *Nature (Lond.)*, **183**, 1373.
INGRAM, V. M. (1958). *Biochim. biophys. Acta*, **28**, 539.
INGRAM, V. M. (1959). *Nature (Lond.)*, **183**, 1795.
ITANO, H. A. (1956). *Ann. Rev. Biochem.*, **25**, 331.
JONES, R. T., SCHROEDER, W. A., BALOG, J. E., and VINOGRAD, J. R. (1959), *J. Amer. chem. Soc.*, **81**, 3161.
JONES, R. T., SCHROEDER, W. A., and VINOGRAD, J. R. (1959). *J. Amer. chem. Soc.*, **81**, in press.
PAULING, L. (1955). *Harvey Lect.*, **41**, 216.
PAULING, L., ITANO, H. A., SINGER, S. J., and WELLS, I. C. (1949). *Science*, **110**, 543.
RHINESMITH, H. S., SCHROEDER, W. A., and PAULING, L. (1957). *J. Amer. chem. Soc.*, **79**, 609.
RIGAS, D. A., KOLER, R. D., and OSGOOD, E. E. (1956). *J. Lab. clin. Med.*, **47**, 51.
SCHROEDER, W. A., KAY, L. M., and WELLS, I. C. (1950). *J. biol. Chem.*, **187**, 221.
SCHWARTZ, H. C., SPAET, T. H., ZUELZER, W. W., NEEL, J. V., ROBINSON, A. R., and KAUFMAN, S. F. (1957). *Blood*, **12**, 238.
SMITH, E. W., and TORBET, J. V. (1958). *Bull. Johns Hopk. Hosp.*, **102**, 38.
VINOGRAD, J. R., HUTCHINSON, W. D., and SCHROEDER, W. A. (1959). *J. Amer. chem. Soc.*, **81**, 3168.